

Serial No. 10/729,889
Wakefield, et al.
Filed December 5, 2003

RECEIVED
CENTRAL FAX CENTER

NOV 19 2006

Response to Office Action
Examiner: Robert M. Timblin
Group Art Unit: 2167

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

INVENTORS: Todd D. Wakefield and David L. Bean

TITLE: Methods and Products for Integrating Mixed format Data Including
The Extraction of Relational Facts from Free Text

FILING DATE: December 5, 2003 **EXAMINER NAME:** Robert M. Timblin

SERIAL NO.: 10/729,889 **GROUP ART UNIT:** 2167

DOCKET: 5133 P

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REMARKS

A terminal disclaimer is attached to overcome the double patenting rejection. Claims 4 and 21 have been amended to overcome the objection.

Section 102. The Examiner rejected claims 1-32 in light of Gaizauskas. Gaizauskas pertains to extracting information from natural language text and recording that information in a structured representation (template). Gaizauskas does not, however, disclose the steps performed by Applicants' claimed invention. An overview of those steps is below, followed by a more detailed treatment of Gaizauskas.

accessing a feed of data records, said data records including both ef-structured data and unstructured data;

the unstructured data of a particular data record including free text related to the structured data

Serial No. 10/729,889
Wakefield, et al.
Filed December 5, 2003

Response to Office Action
Examiner: Robert M. Timblin
Group Art Unit: 2167

of that data record;

extracting relational facts from the free text, said extracting step being performed using linguistic information from the free text;

producing a set of construed data from said unstructured data, each construed datum containing at least one relational fact extracted from free text,

each construed datum being relatable to the structured data of the data record in which said free text was found; and

integrating the construed data with the particular structured data to which the construed data relates.

In particular, Applicant has claimed accessing data records that contain BOTH structured and unstructured data. Gaizauskas accesses ONLY unstructured data. Applicant requests the Examiner to review the sections of Gaizauskas cited in the office action to confirm the fact that Gaizauskas does not access any data records that contain a combination of unstructured and structured data. Therefore Gaizauskas does not disclose this element found in Applicant's claims: " (Claim 1). The other independent claims have similar but slightly different corresponding elements.

Applicant has also claimed the step of extracting relational facts from the free text contained in the data records that have both structured and unstructured data. Gaizauskas discloses only use of unstructured data, therefore Gaizauskas does not disclose "extracting relational facts from the free text, said extracting step being performed using linguistic information from the free text" as claimed by Applicant.

Applicant has also claimed the step of producing construed data with relational facts from the unstructured data and then relating those facts to the structured data through an integration process.. Gaizauskas discloses only accessing unstructured data to fill his templates and perform his further work with the unstructured data alone. He does not produce relational facts from unstructured data and then relate them to structured data found in the same data record as the unstructured data. Therefore Gaizauskas does not disclose:

"producing a set of construed data from said unstructured data, each construed datum containing at

Serial No. 10/729,889
Wakefield, et al.
Filed December 5, 2003

Response to Office Action
Examiner: Robert M. Timblin
Group Art Unit: 2167

least one relational fact extracted from free text,

each construed datum being relatable to the structured data of the data record in which said free text was found".

Applicant has also claimed the step of integrating the construed data with the structured data to create a useful result. Gaizauskus only accesses unstructured data and performs his work from there. He does not begin with data records that contain both structured and unstructured (free text) data mixed together, so he has no way of relating his results back to the structured data of the same data record that the free text came from. Consequently, Gaizauskus does not anticipate Applicant's claims.

Gaizauskus also does not disclose using linguistic information within, relying instead on formatting of, free text in order to create relational facts that relate the free text to structured data in the same data record in which the free text was found. As Gaizauskus does not use linguistic information in this manner, Gaizauskus does not anticipate Applicant's patent claims.

Therefore, Gaizauskus does not disclose the claimed invention for any of the pending claims under section 102 of the Patent Act.

Further, Applicant wishes to explain the teachings of the Gaizauskus reference. The information extraction work described in Gaizauskus makes no reference to combining extracted information with structured data. From the Gaizauskus introduction, "IE may be seen as the activity of populating a structured information source (or database) from an unstructured, or free text, information source." Further, when Gaizauskus does discuss combining IE with another technology, it is with information retrieval (IR), another text-focused task that does not address structured data. Applicant provides a function of integrating relational facts from unstructured data back to the structured data in the same data record in which the unstructured data was found. Gaizauskus does not disclose this integration function at all. From the Gaizauskus introduction, "The contrast between the aims of IE and IR systems can be summed up as: IR retrieves relevant documents from collections, IE extracts relevant information from

Serial No. 10/729,889
Wakefield, et al.
Filed December 5, 2003

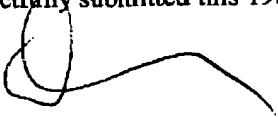
Response to Office Action
Examiner: Robert M. Timblin
Group Art Unit: 2167

documents. The two techniques are therefore complementary, and their use in combination has the potential to create powerful new tools in text processing." Applicant's invention is not about combining IR and IE – it is about combining a particular style of IE with structured data to provide the foundation for a new type of data analysis, as recited in the claims.

The template-filling style of IE described in Gaizauskus is a targeted type of extraction, i.e. the template describes a set of predefined participants or attributes of an event to be populated by the extraction engine. For example, in a kidnapping event, the kidnapping template would have slots for the victim, the perpetrator, the location, and the time of the kidnapping. In one embodiment of Applicants' invention, Applicants describe a system that records all participants of all actions, in a generic form. So, for a kidnapping event, the invention would extract the actor and object of the kidnapping action, but it would also extract the actors and objects of all other events in the text. These events might include physical descriptions of the victim, a list of previous activities by the perpetrator, and a historical recounting of terrorist activities in the area. Unless every possible participant in every possible event has been encoded in a set of templates of the prior art, the targeted approach that Gaizauskus discusses will not cover the exhaustive functions performed by Applicants invention. The invention includes a style of IE that is fundamentally distinct from traditional IE technologies at work only in a targeted fashion.

With respect to specific statements about Gaizauskus made by the Examiner, Applicants believe that Gaizauskus does not make the full disclosure alleged. A comparison of Gaizauskus to the claims as performed above shows the novelty of Applicants' invention. Withdrawal of the rejection is requested.

Respectfully submitted this 19th day of November, 2006.



Daniel P. McCarthy
Reg. No. 36,600
P.O. Box 71550
Salt Lake City, UT 84171-0550
Tel 801 661 8998

Serial No. 10/729,889
Wakefield, et al.
Filed December 5, 2003

Response to Office Action
Examiner: Robert M. Timblin
Group Art Unit: 2167

Fax 435 615 9669
Email: dmccart@xmission.com